

JP09218881

**ADDITIONAL RETRIEVAL WORD CANDIDATE PRESENTING METHOD,
DOCUMENT RETRIEVING METHOD AND DEVICES THEREFOR**

NIPPON TELEGR & TELEPH CORP <NTT>

Inventor(s): INOUE TAKASHI ; TANAKA KAZUO ; TAKESHITA ATSUSHI

Application No. 08048385 , Filed 19960209 , Published 19970819

Abstract:

PROBLEM TO BE SOLVED: To obtain a proper retrieval result by a reretrieval by picking up the word existing in the document obtained by the retrieval for which a prescribed retrieval condition is used and presenting the picked up word as the candidate of a new retrieval word to be added at the retrieval of the next time.

SOLUTION: A memory 11 for retrieval word is a storage part storing a prescribed retrieval word. The prescribed retrieval word is stored to be taken out when the prescribed retrieval word is made AND or OR with an additional retrieval word in the second and succeeding retrievals. A control part 10 is a word pickup means picking up the word existing in the document obtained by the retrieval for which a prescribed retrieval condition is used. An additional retrieval candidate presenting means 20 presents the picked up word as the candidate of a new retrieval word to be added at the retrieval of the next time. Thus, a user can easily acquire truly necessary information in short time.

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-21881

(43)公開日 平成9年(1997)8月19日

(51)Int.Cl.⁶
G 0 6 F 17/30

識別記号 庁内整理番号

F I
G 0 6 F 15/403

技術表示箇所
3 4 0 Z
3 2 0 B

審査請求 未請求 請求項の数9 FD (全16頁)

(21)出願番号 特願平8-48385

(22)出願日 平成8年(1996)2月9日

(71)出願人 000004226
日本電信電話株式会社
東京都新宿区西新宿三丁目19番2号
(72)発明者 井上 孝史
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(72)発明者 田中 一男
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(72)発明者 竹下 敏
東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内
(74)代理人 弁理士 川久保 新一

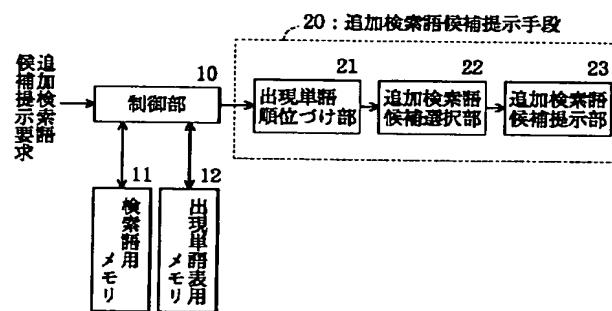
(54)【発明の名称】 追加検索語候補提示方法、文書検索方法およびそれらの装置

(57)【要約】

【課題】 再検索によって適切な検索結果を得ることができ、したがって、ユーザは、本当に必要な情報を短時間にしかも容易に取得することができる追加検索語候補提示方法、文書検索方法およびそれらの装置を提供することを目的とするものである。

【解決手段】 所定の文書をデータベースに予め登録し、所定の検索条件に関連する文書を上記データベースから取り出す場合、所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップし、次回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示するものである。

PS1：追加検索語候補提示装置
(第1の実施例)



【特許請求の範囲】

【請求項 1】 所定の文書をデータベースに登録し、所定の検索条件に関連する文書を上記データベースから取り出す文書検索方法において、

所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップする単語ピックアップ段階と；今回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示する追加検索語候補提示段階と；を有することを特徴とする追加検索語候補提示方法。

【請求項 2】 所定の文書をデータベースに登録し、所定の検索条件に関連する文書を上記データベースから取り出す文書検索方法において、

所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップする単語ピックアップ段階と；今回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示する追加検索語候補提示段階と；上記提示された追加検索語候補から所望の追加検索語を選択する追加検索語選択段階と；上記選択された追加検索語と上記所定の検索条件とを用いて、再び検索を行う再検索段階と；を有することを特徴とする文書検索方法。

【請求項 3】 所定の文書をデータベースに登録し、所定の検索条件に関連する文書を上記データベースから取り出す文書検索装置において、

所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップする単語ピックアップ手段と；今回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示する追加検索語候補提示手段と；を有することを特徴とする追加検索語候補提示装置。

【請求項 4】 所定の文書をデータベースに登録し、所定の検索条件に関連する文書を上記データベースから取り出す文書検索装置において、

所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップする単語ピックアップ手段と；今回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示する追加検索語候補提示手段と；上記提示された追加検索語候補から所望の追加検索語を選択する追加検索語選択手段と；上記選択された追加検索語と上記所定の検索条件とを用いて、再び検索を行う再検索手段と；を有することを特徴とする文書検索装置。

【請求項 5】 請求項 4において、

上記単語ピックアップ手段は、所定の検索条件に基づく検索によってヒットした所定の文書を得た後に、上記ヒットした所定の文書中の 1 つの文書に存在する所定の単語と、上記ヒットした所定の文書中の他の文書に存在する単語とがともに存在する文書の数を検出する手段であることを特徴とする文書検索装置。

【請求項 6】 請求項 4において、

上記単語ピックアップ手段は、データベース中の全文書に存在する単語をピックアップし、このピックアップされた単語中の所定の単語と、このピックアップされた単語中の所定の単語とともに同一の文書に存在する共起単語と、上記共起単語が上記所定の単語とともに存在する文書の数である共起文書数とが対応する共起表を、検索前に作成し、検索してヒットした文書中に存在する各単語について、対応する共起単語と共起文書数とを上記共起表から取り出す手段であることを特徴とする文書検索装置。

【請求項 7】 請求項 4 または請求項 5において、追加検索語候補提示手段は、上記ピックアップされた単語と、このピックアップされた単語が存在する文書の数とを対応させて提示する手段であることを特徴とする文書検索装置。

【請求項 8】 請求項 4～7 のいずれか 1 項において、追加検索語候補提示手段は、上記ピックアップされた単語が存在する文書の数に応じた所定の基準によって、上記ピックアップされた単語を順位付けて提示する手段であることを特徴とする文書検索装置。

【請求項 9】 請求項 4において、上記再検索手段は、上記選択された追加検索語と上記所定の検索語とを、AND 条件または OR 条件で検索する手段であることを特徴とする文書検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、所定の文書をデータベースに登録し、所定の検索語が含まれる文書等を上記データベースから取り出す場合における追加検索語候補提示方法、文書検索方法およびそれらの装置に関するものである。

【0002】

【従来の技術】全文検索とは、1 つの文書の全体を登録対象として、複数の文書をデータベースに予め登録し、ユーザが与えた検索条件に関連する文書をそのデータベースから取り出す技術である。検索条件は、たとえば「通信」のように 1 つの検索語だけに関連する文書、「通信 AND 計算機」のように「通信」と「計算機」との両方の検索語に関連する文書、「通信 OR 計算機」のようにいづれかの検索語に関連する文書も受諾されることが多い。

【0003】ここで、「所定の検索条件に関連する文書」とは、「所定の検索語が含まれる文書」とほぼ同義であり、「所定の検索語と同一の単語が含まれる文書」以外に、「所定の検索語と同義の単語が含まれる文書」、「所定の検索語と送りがなみが異なる単語が含まれる文書」も含まれる。

【0004】図 17 は、従来の一般的な検索方法を示すフローチャートである。

【0005】図17に示すフローチャートにおいて、まず、ユーザが1つの検索語を与えて検索し(S1)、その検索結果がディスプレイに表示され(S2)、この検索結果に応じて、ユーザが新たな検索語を考え、この考えた新たな検索語を追加することによって検索条件を変更し(S3)、再度検索する(S4)。

【0006】たとえば、最初「通信」という語を検索条件として検索したときに、この検索の結果、希望する文書数よりも多くの文書が検索された場合、「通信AND計算機」等のように検索条件を変更して検索条件を絞り込む。

【0007】

【発明が解決しようとする課題】従来の全文検索方法においては、最初の検索結果に対して、ユーザが、AND条件で検索語を追加することによって検索条件を変更する場合、どの検索語を新たに追加すればよいかという明確な指針が存在しない。したがって、ユーザは、経験と直感とに頼って新たな検索語を決め、この決められた検索語を追加して再検索するが、このようにして再検索しても、以前の検索結果とほとんど変化がなかったり、または逆に、必要以上に検索結果が少なくなる（典型的には空になる）等、適切な結果が得られないことが多い。

【0008】従来方法においては、検索条件の変更が試行錯誤で何度も行なわれることになり、結局、検索効率が悪いという問題がある。この問題は、文書の一部を検索対象とした場合、文書のキーワードのみを検索対象とした場合のように、全文検索以外の検索の場合も同様に生じる問題である。

【0009】本発明は、再検索によって適切な検索結果を得ることができ、したがって、ユーザは、本当に必要な情報を短時間にしかも容易に取得することができる追加検索語候補提示方法、文書検索方法およびそれらの装置を提供することを目的とするものである。

【0010】

【課題を解決するための手段】本発明は、所定の文書をデータベースに予め登録し、所定の検索条件に関連する文書を上記データベースから取り出す場合、所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップし、次回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示するものである。

【0011】

【発明の実施の形態および実施例】図1は、本発明の一実施例である追加検索語候補提示装置PS1を示すブロック図である。

【0012】追加検索語候補提示装置PS1は、制御部10と、検索語用メモリ11と、出現単語表用メモリ12と、追加検索語候補提示手段20とを有する。追加検索語候補提示手段20は、出現単語順位付け部21と、追加検索語候補選択部22と、追加検索語候補提示部250

3とを有する。

【0013】追加検索語候補提示装置PS1は、全文検索の対象になる文書全体をデータベースに登録し、所定の検索条件に関連する文書を上記データベースから取り出す文書検索装置において、追加すべき検索語の候補を提示する装置である。ここで、「所定の検索条件に関連する文書」は、「所定の検索語と同一の単語が含まれる文書」、「所定の検索語と同義の単語が含まれる文書」、「所定の検索語と送りがなが異なる単語が含まれる文書」である。また、上記の場合、所定の検索語とは、1つの検索語または複数の検索語のことである。

【0014】検索語用メモリ11は、所定の検索語を記憶する記憶部であり、2回目以降の検索において、上記所定の検索語を追加検索語とANDまたはORする場合に、その所定の検索語を取り出すために記憶させるものである。出現単語表用メモリ12は、図2に示す出現単語表T1を記憶する部分である。

【0015】制御部10は、所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップする単語ピックアップ手段の例である。追加検索語候補提示手段20は、次回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示する追加検索語候補提示手段の例である。

【0016】図2は、上記実施例における出現単語表T1の例を示す図である。

【0017】出現単語表T1は、ヒットした所定の文書に存在する全ての単語をピックアップし、このピックアップされた各単語と、このピックアップされた各単語が存在する文書（ヒットした所定の文書に限る）の数とを対応させた表である。

【0018】つまり、所定の検索語に基づく検索によってヒットした文書中に、「ネットワーク」、「システム」等の単語が存在し、そのうちの「ネットワーク」という単語が含まれるヒット文書が10個あり、「システム」という単語が含まれるヒット文書が4つ存在している例が図2に示されている。

【0019】なお、出現単語表T1に表示される単語は、助詞、助動詞等の付属語を含まない自立語である。

【0020】追加検索語候補提示手段20は、所定の検索語に対応する単語を追加検索語候補として提示する手段であり、出現単語順位付け部21と、追加検索語候補選択部22と、追加検索語候補提示部23とを有する。

【0021】出現単語順位付け部21は、出現単語を所定の方法で順位付けする部分であり、追加検索語候補選択部22は、この順位付けされた出現単語を所定の順序で所定数選択する部分であり、追加検索語候補提示部23は、この追加検索語候補選択部22が選択した出現単語を追加検索語候補として提示する部分である。

【0022】なお、追加検索語候補を提示する場合、その追加検索語候補を追加したときに検索結果が空になら

ないという条件（第1の条件）と、その追加検索語候補を追加したときの検索結果の文書数が適切に減少するという条件（第2の条件）とを満たすものとする。

【0023】図3は、上記実施例において出現単語を順位付けする場合に使用する関数の一例をグラフ化したものである。

【0024】図3に示すグラフにおいて、その横軸は、（所定の出現単語に関する出現文書数）／（検索された文書の数）であり、0～1の値を取るものであり、その縦軸は、追加検索語としての出現単語の評価を示す点数であり、この点数が高い程、その出現単語の上記順位が高くなるとする。

【0025】図4は、上記実施例において、順位付けされた出現単語の例を示す図である。図4に示す例において、出現単語「ファイル」、「システム」、「通信」、「処理」、「ネットワーク」、……の順位は、それぞれ、1位、2位、3位、4位、5位、……である。

【0026】図5は、上記実施例において、追加検索語候補選択部22によって選択された追加検索語候補の例を示す図である。図5に示す例は、出現単語順位付け部21によって順位付けられた追加検索語候補のうち、順位1～4位のものが選択されている。

【0027】次に、上記実施例の動作について説明する。

【0028】図6は、上記第1の実施例の動作を示すフローチャートである。

【0029】まず、1つの単語が最初の検索語として入力され（S1）、この検索語に対する検索が終わり、この検索結果が表示されたが（S2）、ヒットした文書数が多いので、ユーザはその結果に満足せず、新たな検索語をANDで追加したいと考えたとする。ここで、従来例では、ユーザが任意に思いつく単語を検索語として追加することになるが、上記実施例においては、ユーザから要求があると有効な追加検索語の候補が提示され（S10）（または、ユーザからの要求がなくても自動的に有効な追加検索語の候補が提示され）、ユーザはその中から追加すべき検索語を任意に選択し（S20）、選択した検索語を追加し（S3）、再検索する（S4）。

【0030】図7は、上記実施例における追加検索語候補提示動作（S10）を具体的に示すフローチャートである。

【0031】図7に示すフローチャートにおいて、ユーザからの追加検索語候補提示要求を制御部10が受けると（S11）、制御部10が出現単語表T1を作成する（S12）。そして、その出現単語を取り出し、順位付けする（S13）。順位づけの方法としては種々の方法を考えられるが、出現文書数の少ないものから順位づける方法（第1の方法）や、出現文書数が中程度のものを絞り込むことがより有効であるとして、図3に示すグラフの関数に応じて、点数づけし（評価し）、その点数

の高いものから順位づけする方法（第2の方法）等が考えられる。

【0032】このように順位づけされた出現単語の中から、所定数の出現単語を絞って追加検索語候補とするような追加検索語候補選択を行なう（S14）。この場合、適当な閾値を定め、この定められた閾値よりも低い順位の出現単語を、追加検索語候補から除去する。

【0033】このようにして残った出現単語を追加検索語候補として、出現文書数とともにユーザに提示する（S15）。ユーザは、この中から追加検索語を任意に選び出し、検索条件にAND条件として追加し（S3）、再検索を行なう（S4）。

【0034】図8は、上記実施例における出現単語表T1の作り方（S12）を示すフローチャートである。

【0035】まず、空の出現単語表T1を作成し（S21）、直前に設定した検索語（検索条件）でヒットした文書の集合を文書群Sdとし（S22）、ヒット文書群Sdから1つの文書Dを抜き出し（S24）、文書Dに出現する（存在する）単語の集合を単語群Swとする（S25）。ただし、単語群Sw中に同一の単語は1つのみ含むとする。そして、単語群Swから1つの単語Wを抜き出し（S27）、この単語Wが出現単語表T1の出現単語欄に存在していなければ（S28）、出現単語表T1の出現単語欄に単語Wを追加し、対応する出現文書数の欄を「1」にする（S29）。単語Wが出現単語表T1の出現単語欄に存在していれば（S28）、2つ目以降の文書が検出されたのであるから、出現単語表T1の出現単語欄に既に記載されている単語Wに対応する出現文書数の欄の数字を1インクリメントする（S30）。

【0036】そして、単語群Swから当該単語Wを除去し（S31）、単語群Swから次の単語Wを抜き出し（S26、S27）、上記動作（S28～S31）を実行し、これらを繰り返し、単語群Swに単語Wが無くなれば（S26）、ヒット文書群Sdから当該文書Dを取り除く（S32）。その後、ヒット文書群Sdから次の文書Dを抜き出し（S23、S24）、上記動作（S25～S32）を実行し、これらを繰り返し、ヒット文書群Sdに文書Dが無くなれば（S23）、リターンする。

【0037】次に、上記実施例をより具体的に説明する。

【0038】まず、ユーザは「計算機」という単語を最初の検索語として検索を行ない、その検索された文書数が当初の希望数よりも多かったとし、AND条件で新たな検索語を追加するために、追加検索語候補提示要求を出したとする。このときに、図2に示す出現単語表T1が制御部10によって与えられ、「出現単語順位づけ方法」として、「出現文書数の少ないものから順位づける方法」を採用することとし、追加検索語候補選択部2

2の閾値として「4」を用いる（順位第4位までを抜き出す）ものとする。

【0039】ここで、要求を受けた制御部10は、出現単語表T1を調べ、結果として図5に示す順位づけされた出現単語を得る。

【0040】最後に、追加検索語候補選択部22において、図5に示すように、順位の上位4個の出現単語が残り、この上位4個の出現単語を追加検索語候補としてユーザーに提示する。この場合、その候補を表示装置等に表示することによって提示する。ユーザーは、提示された4個の追加検索語候補の中から適当なものを選び、この選ばれた追加検索語を検索条件に追加し、再検索を行なう。

【0041】上記実施例によれば、所定の検索語による検索結果に応じて、AND条件で追加すべき有効な追加検索語の候補を、ユーザに提示するようにしているので、ユーザ自身の勘に頼ることなく、ユーザがその中から追加検索語を容易に選択でき、また、このようにして選択された追加検索語を追加して検索するので、より適切な文書を得ることができる。

【0042】なお、追加検索語候補提示手段20が、追加検索語候補提示部23のみで構成されているものであるとしてもよく、この場合、追加検索語候補提示手段20は、所定の検索語に対応する出現単語を追加検索語候補として提示する手段であり、順位付け、追加検索語の選択を行わない。このようにしても、従来方法と比較すると、必要な情報を短時間にしかも容易に取得することができる。

【0043】また、追加検索語候補提示手段20が、出現単語順位付け部21と、追加検索語候補提示部23とで構成され、追加検索語候補選択部22が削除されたものであるとしてもよく、この場合、追加検索語候補提示手段20は、出現単語を所定の方法で順位付けする出現単語順位付け部と、この順位付けされた出現単語を上記順位とともに追加検索語候補として提示する追加検索語候補提示部とで構成されている。このようにすれば、追加検索語候補提示手段20が追加検索語候補提示部23のみで構成されている場合よりも、2回目の検索が容易になる。

【0044】上記実施例において、追加検索語選択22は、提示された追加検索語候補から所望の追加検索語を選択する追加検索語選択手段の例であり、制御部10は、選択された追加検索語と所定の検索語とを用いて、再び検索を行う再検索手段の例である。

【0045】また、制御部10は、単語ピックアップ手段の例であるが、この単語ピックアップ手段は、所定の検索条件に基づく検索によってヒットした所定の文書を得た後に、上記ヒットした所定の文書中の1つの文書に存在する所定の単語と、上記ヒットした所定の文書中の他の文書に存在する単語とがともに存在する文書の数を

検出する手段である。

【0046】追加検索語候補提示装置PS1においては、所定の検索語に基づく検索によって所定の文書を得ており、1つの検索語を使用して検索しているが、2つ以上の検索語を同時に使用して検索するようにもよく、つまり、所定の検索条件で検索するようにもよい。

【0047】図9は、本発明の他の実施例である追加検索語候補提示装置PS2を示すブロック図である。

【0048】追加検索語候補提示装置PS2は、制御部30と、検索語用メモリ31と、共起表用メモリ32と、追加検索語候補提示手段40とを有する。追加検索語候補提示手段40は、共起単語順位付け部41と、追加検索語候補選択部42と、追加検索語候補提示部43とを有する。

【0049】追加検索語候補提示装置PS2は、全文検索の対象になる文書全体をデータベースに登録し、所定の検索条件に関連する文書を上記データベースから取り出す文書検索装置であり、基本的には、追加検索語候補提示装置PSと同じであるが、出現単語表T1の代わりに共起表T2を使用し、制御部10の代わりに制御部30を使用し、追加検索語候補提示手段20の代わりに、追加検索語候補提示手段40を使用している点が異なる。

【0050】また、共起表T2は、原則として、データベース完成時に1回のみ作れば足りるものであり、検索する度に作る必要はない（データベースを更新した場合には共起表も更新する）。制御部30は、基本的には、制御部10と同じであるが、出現単語表T1の代わりに共起表T2を使用することに応じて動作が多少異なる。また、追加検索語候補提示手段40も、基本的には、追加検索語候補提示手段20と同じであるが、出現単語表T1の代わりに共起表T2を使用することに応じて、その構成、動作が多少異なる。検索語用メモリ31は、最初の検索語を記憶する記憶部である。

【0051】制御部30は、単語ピックアップ手段の例であり、この単語ピックアップ手段は、データベース中の全文書に存在する単語をピックアップし、このピックアップされた単語中の1つの単語と、このピックアップされた単語中の1つの単語とともに同一の文書に存在する共起単語と、共起文書数とが対応する共起表を、検索前に作成し、検索してヒットした文書中に存在する各単語について、上記共起表から対応する共起単語と共起文書数とを取り出す手段の例である。

【0052】さらに、共起単語順位付け部41は、ピックアップされた単語が存在する文書の数に応じた所定の基準によって、上記ピックアップされた単語を順位付けて提示する手段の例であり、追加検索語候補提示部43は、ピックアップされた単語と、このピックアップされた単語が存在する文書の数とを対応させて提示する手段

の例である。

【0053】ところで、「共起単語」は、データベース中の全文書に存在する単語をピックアップし、このピックアップされた単語中の1つの単語と、このピックアップされた単語中の1つの単語とともに同一の文書に存在する単語である。また、「共起文書数」は、ピックアップされた単語中の1つの単語とともに共起単語が存在する文書の数である。「共起表」は、ピックアップされた単語中の1つの単語と、その共起単語と、その共起文書数とが対応している表である。

【0054】図10は、上記第2の実施例における共起表T2の例を示す図である。

【0055】図10に示す共起表T2において、各行の左端に示す「単語の欄」には、データベースに登録されている文書中の単語が記載され、「単語の欄」の右に記載されている「共起単語の欄」には、「単語の欄」に記載されている単語と共起する単語（同一文書中に存在する単語）が列挙されている。なお、「共起単語」の欄に記載されている括弧中の数字は、共起単語が記載されている文書の数である。また、「共起単語」として採用される単語は、助詞、助動詞等の付属語を含まない自立語である。

【0056】つまり、データベースに登録されている文書中には、「ファイル」、「計算機」、「処理」等の単語が含まれ、たとえば「計算機」という単語と共に起する単語（共起単語）、すなわち、「計算機」という単語が含まれる文書中に含まれる単語は、図10の例によれば、「通信」、「ネットワーク」、「処理」、「ファイル」等である。この場合、「通信」という単語が含まれる文書はデータベースに3つ登録され、「ネットワーク」という単語が含まれる文書はデータベースに10個登録されている。

【0057】制御部30は、データベース完成時に共起表T2を作成する部分であり、この作成した共起表を共起表用メモリ32に記憶し、この共起表T2から、所定の検索語に対応する共起単語を追加検索語として取得する部分である。たとえば、所定の検索語を「計算機」とした場合、「通信」、「ネットワーク」、「処理」、「ファイル」等の共起単語を追加検索語として取得する。

【0058】共起単語順位付け部41は、共起単語を所定の方法で順位付けする部分であり、追加検索語候補選択部42は、この順位付けされた共起単語を所定の順序で所定数選択する部分であり、追加検索語候補提示部43は、この追加検索語候補選択部が選択した上記共起単語を追加検索語候補として提示する部分である。

【0059】図11は、上記第2の実施例において共起単語を順位付けする場合に使用する関数の一例をグラフ化したものである。

【0060】図11に示すグラフにおいて、その横軸

は、（最初の検索語の共起単語が含まれる文書の数）／（最初の検索によって得られた文書の数）であり、0～1の値を取るものであり、その縦軸は、追加検索語としての共起単語の評価を示す点数であり、この点数が高い程、その共起単語の上記順位が高くなるとする。

【0061】図12は、上記第2の実施例において、順位付けされた共起単語の例を示す図である。図12に示す例において、共起単語「ファイル」、「システム」、「通信」、「処理」、「ネットワーク」、……の順位は、それぞれ、1位、2位、3位、4位、5位、……である。

【0062】図13は、上記第2の実施例において、追加検索語候補選択部42によって選択された追加検索語候補の例を示す図である。図13に示す例は、共起単語順位付け部41によって順位付けられた追加検索語候補のうち、順位1～4位のものが選択されている。

【0063】次に、上記第2の実施例の動作について説明する。

【0064】図14は、上記第2の実施例の動作を示すフローチャートである。

【0065】まず、データベース完成時に共起表T2を作成しておく（S0）。そして、検索時に、1つの単語が最初の検索語として入力され（S1）、この検索語に対する検索が終わり、この検索結果が表示されたが（S2）、ヒットした文書数が多いので、ユーザはその結果に満足せず、新たな検索語をANDで追加したいと考えたとする。ここで、従来例では、ユーザが任意に思いつく単語を検索語として追加することになるが、上記第2の実施例においては、ユーザから要求があると有効な追加検索語の候補が提示され（S10a）（または、ユーザからの要求がなくても自動的に有効な追加検索語の候補が提示され）、ユーザはその中から追加すべき検索語を選択し（S20a）、選択した検索語を追加し（S3）、再検索する（S4）。

【0066】図15は、上記第2の実施例における追加検索語候補提示動作（S10a）を具体的に示すフローチャートである。

【0067】図15に示すフローチャートにおいて、ユーザが追加検索語候補提示を制御部30に要求すると（S11）、最初の検索語を検索語用メモリ31から取り出し、この最初の検索語と共に起する共起単語を、共起表T2の対応する行から取り出し、このときに、共起単語とともに共起する文書の数（共起単語が含まれる文書の数）も取り出す（S12a）。

【0068】このときに取り出した共起単語が、追加検索語の初期候補になる。つまり、最初の検索語とその取り出した共起単語とをAND条件で追加検索した場合に、その検索結果として得られる文書は、最初の検索語と追加された共起単語との両方が同時に含まれている文書であるので、第1の条件（その語を追加したときに検

索結果が空にならないという条件) を満たす共起単語は、追加検索語の初期候補になり、しかも、その共起単語は必ず第1の条件を満たす。また、共起表T2中の共起文書の数は、その共起単語を最初の検索語とAND条件で追加検索した場合に検索される文書数である。

【0069】そして、取り出された共起単語に順位づけを行なう(S13a)。順位づけの方法としては種々の方法が考えられるが、共起文書数の少ないものから順位づける方法(第1の方法)や、共起文書数が中程度のものを絞り込むことがより有効であるとして、図11に示すグラフの関数に応じて、点数づけし(評価し)、その点数の高いものから順位づける方法(第2の方法)等が考えられる。

【0070】このように順位づけされた共起単語の中から、所定数の共起単語を絞って追加検索語候補とするような追加検索語候補選択を行なう(S14a)。この場合、適当な閾値を定め、この定められた閾値よりも低い順位の共起単語を、追加検索語候補から除去する。

【0071】このようにして残った共起単語を追加検索語候補として、共起文書数とともにユーザに提示する(S15a)。ユーザは、この中から追加検索語を選び出し、検索条件にAND条件として追加し(S3)、再検索を行なう(S4)。

【0072】図16は、上記第2の実施例における共起表T2の作り方(S0)を示すフローチャートである。

【0073】まず、空の共起表T2を作成し、テキストデータベース中の文書の集合をSdとし(S41)、文書群Sdから1つの文書Dを抜き出し(S43)、文書Dに存在する全ての単語の集合を単語群Swとする(S44)。ただし、単語群Sw中に同一の単語は1つのみ含むとする。そして、単語群Swから1つの単語Wを抜き出し(S46)、この単語Wが共起表T2の単語欄に存在していなければ(S47)、共起表T2の単語欄に単語Wを追加する(S48)。

【0074】そして、文書D中に存在する全ての単語の集合を、共起単語検索用の単語群Scとする(S49)。ただし、単語群Sc中に同一の単語は1つのみ含むとする。その後、単語群Scから1つの単語Wcを抜き出し(S52)、単語Wcが共起表T2の単語Wに対応する共起単語欄に存在しなければ(S53)、共起表T2の単語Wに対応する共起単語の欄に単語Wcを追加し、その共起文書の数を「1」にする(S54)。単語Wcが共起表T2の単語Wに対応する共起単語欄に存在すれば(S53)、単語Wcに対応する共起文書の数を1インクリメントし(S55)、単語群Scから単語Wcを取り除き(S56)、単語群Scから次の単語Wcを抜き出し(S51、S52)、上記動作(S53～S56)を実行し、これらを繰り返す。

【0075】そして、単語群Scに単語Wが無くなれば(S51)、所定単語に関する共起単語を検索するため

の単語としての単語Wを単語群Swから取り除き(S57)、新たな単語について共起単語を検索するために、単語群Swから次の単語Wを抜き出し(S46)、上記動作(S47～S56)を実行し、これらを繰り返す。そして、Swが無くなれば(S45)、文書群Sdから文書Dを取り除き(S58)、文書群Sdから次の文書Dを抜き出し(S43)、上記動作(S44～S56)を繰り返し、文書群Sdに文書Dが無くなれば(S42)、リターンする。

【0076】次に、上記第2の実施例をより具体的に説明する。

【0077】まず、ユーザは「計算機」という単語を最初の検索語として検索を行ない、その検索された文書数が当初の希望数よりも多かったとし、AND条件で新たな検索語を追加するために、追加検索語候補提示要求を出したとする。このときに、共起表用メモリ32に図10に示す共起表T2が記憶されているので、候補提示要求を受けた制御部30は、検索語用メモリ31から最初の検索語である「計算機」を取り出し、共起表T2を調べ、「計算機」の右に書かれている共起単語を取り出す。この例では「通信(5)」、「ネットワーク(10)」、「処理(8)」、「ファイル(3)」、「ソフトウェア(10)」、「システム(4)」等を取り出す。なお、括弧内は共起文書の数である。

【0078】次に、共起単語順位づけ部41において、制御部30で取り出した共起単語の順位づけを行う。順位づけの方法として、「共起文書の少ないものから順位づける方法」を採用しているから、結果として図13に示す順位づけされた共起単語を得る。

【0079】最後に、追加検索語候補選択部42において、図13に示すように、順位の上位4個の共起単語が残り、この上位4個の共起単語を追加検索語候補としてユーザーに提示する。ユーザーは、提示された4個の追加検索語候補の中から適当なものを選び、この選ばれた追加検索語を検索条件に追加し、再検索を行なう。

【0080】上記第2の実施例によれば、所定の検索語による検索結果に応じて、AND条件で追加すべき有効な追加検索語の候補をユーザに提示するようにしているので、ユーザがその中から追加検索語を選んで追加するので、より適切な文書を得ることができ、しかも追加検索語の選択が容易である。さらに、上記第2の実施例によれば、共起表T2がデータベース完成時に作成されるので、検索の度に共起表T2を作成する必要がなく、この作成時間だけ、追加検索語の候補を提示する準備時間が短くなる。

【0081】上記第2の実施例においては、共起表T2における「単語の欄」の1行に1つの単語のみが記載されているが、その1行に2つ以上の単語を記載するようにしてもよい。この場合、共起表における「単語の欄」の1行に2つ以上の単語が記載されている共起表を、デ

ータベース完成時に作成してもよく、また、検索時に作成するようにしてもよい。

【0082】なお、上記各実施例において、再検索する場合、選択された追加検索語と所定の検索語とを、AND条件で検索するようにしてもよく、また、選択された追加検索語と所定の検索語とを、OR条件で検索するようにしてもよい。このようにOR条件で検索できるようすれば、最初の検索語（所定の検索語）による検索がいわゆる絞り込み過ぎていた場合に、より適切な検索を行うことができる。

【0083】また、上記各実施例を方法として把握すると、所定の文書をデータベースに登録し、所定の検索条件に関連する文書を上記データベースから取り出す文書検索方法において、所定の検索条件を用いた検索によって得られた文書中に存在する単語をピックアップする単語ピックアップ段階と、次回の検索時に追加すべき新たな検索語の候補として、上記ピックアップされた単語を提示する追加検索語候補提示段階とを有する追加検索語候補提示方法である。また、この方法において、上記提示された追加検索語候補から所望の追加検索語を選択する追加検索語選択段階と、上記選択された追加検索語と上記所定の検索条件とを用いて、再び検索を行う再検索段階とを追加することによって、文書検索方法を構成するようにしてもよい。

【0084】また、上記各実施例は、文書の一部を検索対象とした場合、文書のキーワードのみを検索対象とした場合等のように、全文検索する場合以外の検索にも適用できるものである。

【0085】

【発明の効果】本発明によれば、再検索によって適切な検索結果を得ることができ、したがって、ユーザは、本当に必要な情報を短時間にしかも容易に取得することができるという効果を奏する。

【図面の簡単な説明】

【図1】本発明の一実施例である追加検索語候補提示装置PS1を示すブロック図である。

【図2】上記実施例における出現単語表T1の例を示す図である。

【図3】上記実施例において出現単語を順位付けする場合に使用する関数の一例をグラフ化したものである。

【図4】上記実施例において、順位付けされた共起単語の例を示す図である。

* 【図5】上記実施例において、追加検索語候補選択部22によって選択された追加検索語候補の例を示す図である。

【図6】上記第1の実施例の動作を示すフローチャートである。

【図7】上記実施例における追加検索語候補提示動作(S10)を具体的に示すフローチャートである。

【図8】上記実施例における出現単語表T1の作り方(S12)を示すフローチャートである。

10 【図9】本発明の他の実施例である追加検索語候補提示装置PS2を示すブロック図である。

【図10】上記第2の実施例における共起表T2の例を示す図である。

【図11】上記第2の実施例において共起単語を順位付けする場合に使用する関数の一例をグラフ化したものである。

【図12】上記第2の実施例において、順位付けされた共起単語の例を示す図である。

20 【図13】上記第2の実施例において、追加検索語候補選択部42によって選択された追加検索語候補の例を示す図である。

【図14】上記第2の実施例の動作を示すフローチャートである。

【図15】上記第2の実施例における追加検索語候補提示動作(S10a)を具体的に示すフローチャートである。

【図16】上記第2の実施例における共起表T2の作り方(S0)を示すフローチャートである。

【図17】従来の一般的な検索方法を示すフローチャートである。

【符号の説明】

PS1、PS2…追加検索語候補提示装置、

10、30…制御部、

11、31…検索語用メモリ、

12…出現単語表用メモリ、

20、40…追加検索語候補提示手段、

21…出現単語順位付け部、

22、42…追加検索語候補選択部、

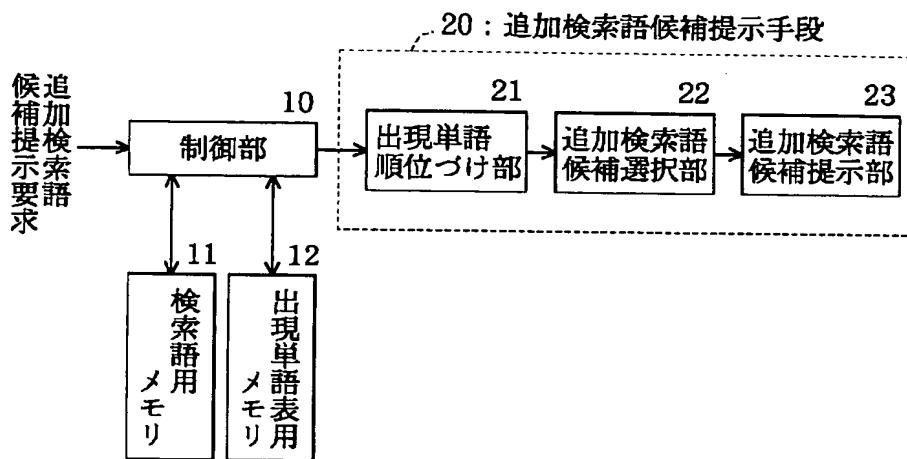
23、43…追加検索語候補提示部、

32…共起表用メモリ、

41…共起単語順位付け部。

【図1】

PS1：追加検索候補提示装置
(第1の実施例)



【図2】

【図5】

選択された追加検索語候補の例

追加検索語候補 (出現文書数)
ファイル (3)
システム (4)
通信 (5)
処理 (8)
⋮

【図1-3】

選択された追加検索語候補の例

追加検索語候補 (共起文書数)
ファイル (3)
システム (4)
通信 (5)
処理 (8)
⋮

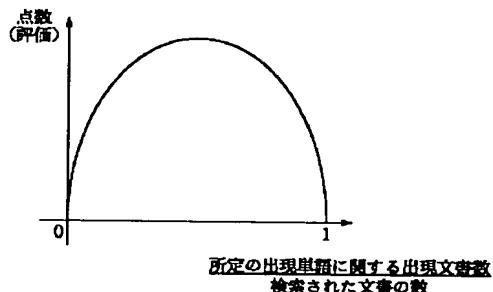
【図3】

出現単語表T1の例

出現単語	出現文書数 (出現単語が存在する文書の数)
ネットワーク	10
システム	4
ファイル	3
処理	8
ソフトウェア	10
⋮	⋮
⋮	⋮

K3718

所定の出現単語を評価するための特性

所定の出現単語に関する出現文書数
検索された文書の数

【図4】

【図10】

順位づけされた出現単語の例

順位	出現単語 (出現文書数)
1	ファイル (3)
2	システム (4)
3	通信 (5)
4	処理 (8)
5	ネットワーク (10)
6	ソフトウェア (10)
⋮	⋮
⋮	⋮

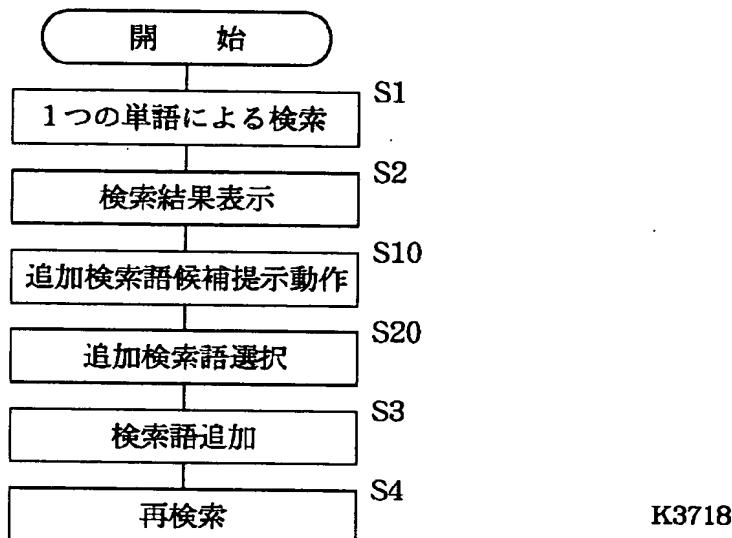
K3718

共起表T2の例

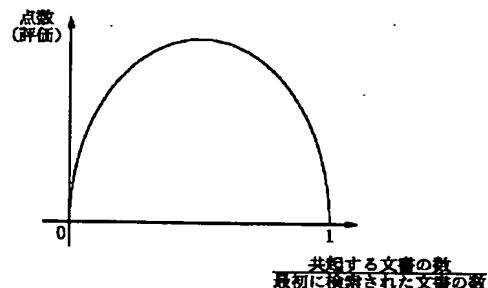
単語	共起単語 (共起文書の数)
ファイル	計算機(3)、事務(22)、OA(17)、処理(10)、共有(8)、……
⋮	⋮
計算機	通信(5)、ネットワーク(10)、処理(8)、事務(3)、ソフトウェア(10)、システム(4)、……
⋮	⋮
処理	文書(12)、計算機(8)、ネットワーク(2)、事務(10)、……
⋮	⋮

K3718

【図6】

第1の実施例の動作

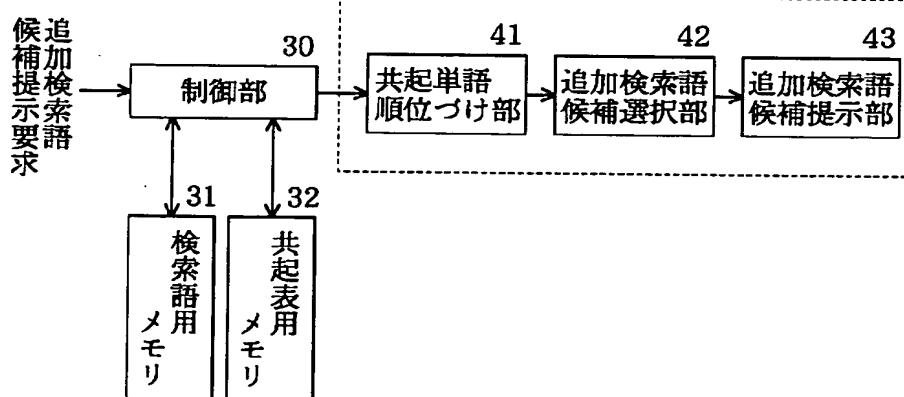
【図11】



【図9】

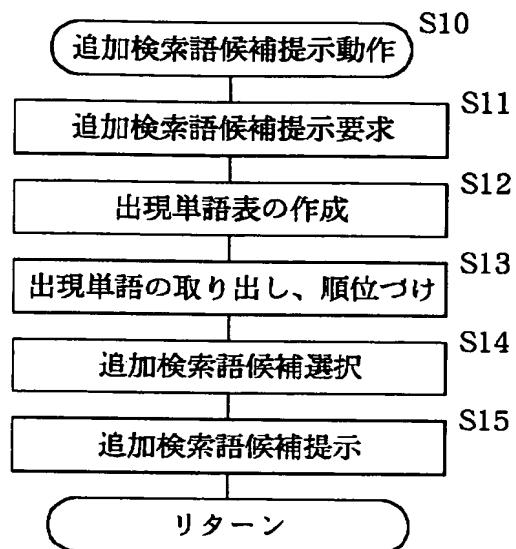
PS2: 追加検索候補提示装置
(第2の実施例)

40: 追加検索語候補提示手段



【図7】

第1の実施例における
追加検索語候補提示動作



【図12】

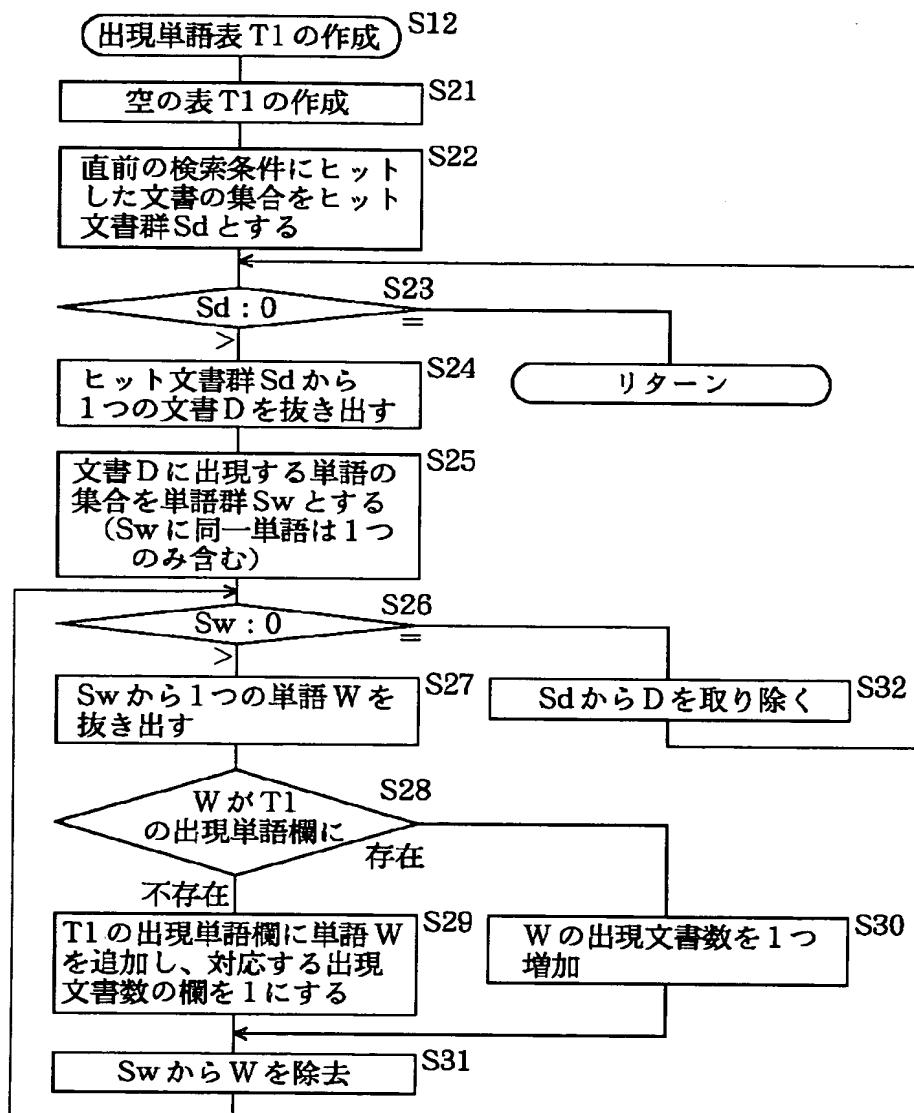
順位づけされた共起単語の例

順位	共起単語（共起文書数）
1	ファイル (3)
2	システム (4)
3	通信 (5)
4	処理 (6)
5	ネットワーク (10)
6	ソフトウェア (10)
⋮	⋮

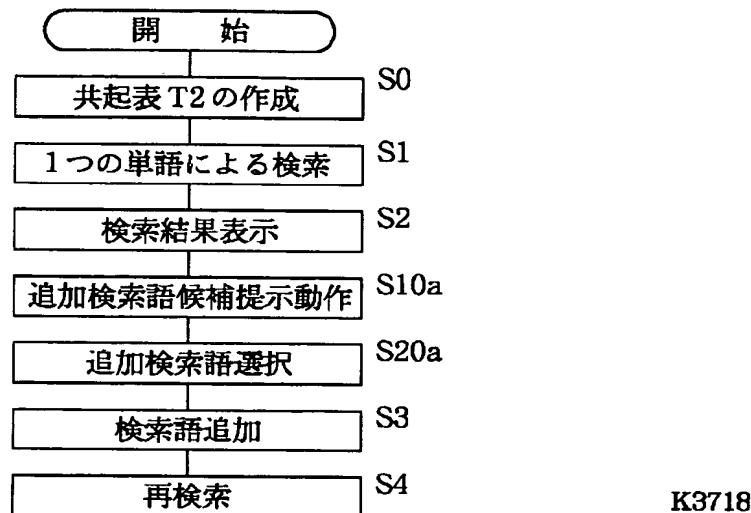
K3718

K3718

【図8】

出現単語表T1の作り方

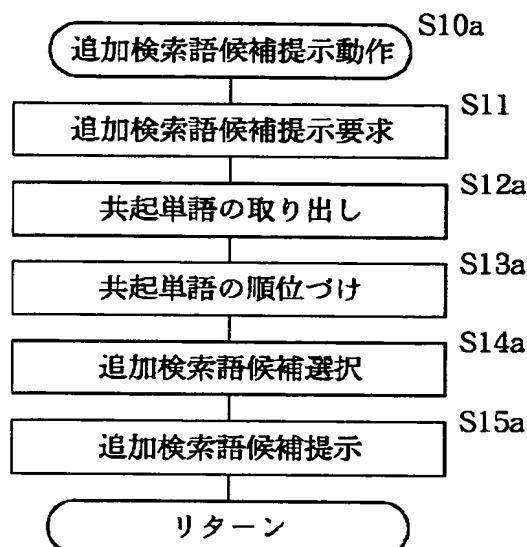
【図14】

実施例の動作

K3718

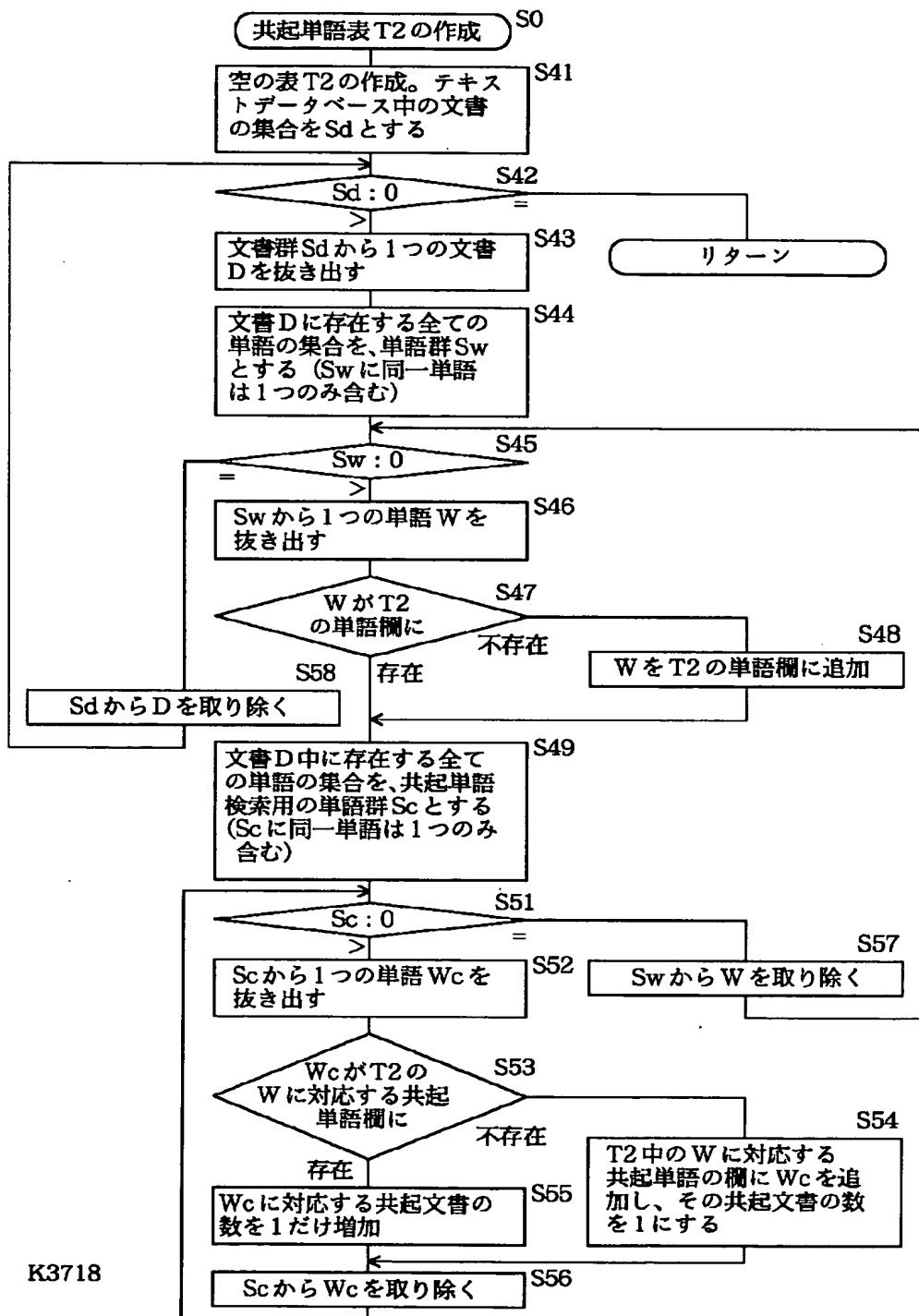
【図15】

第2の実施例における
追加検索語候補提示動作

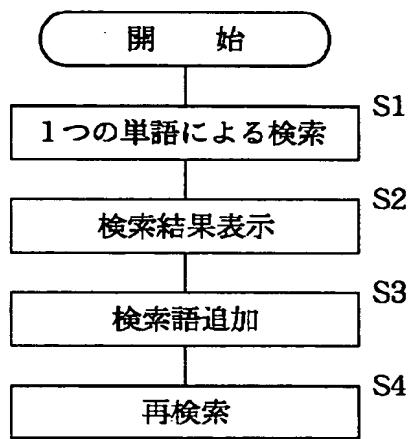


K3718

【図16】

共起表T2の作り方

【図17】

従来例の動作

K3718